Supplementary Materials for "SSFE-Net: Self-Supervised Feature Enhancement for Ultra-Fine-Grained Few-Shot Class Incremental Learning"

Zicheng Pan, Xiaohan Yu, Miaohua Zhang, Yongsheng Gao
School of Engineering and Built Environment
Griffith University, QLD, 4111, Australia
{z.pan; xiaohan.yu; lena.zhang; yongsheng.gao}@griffith.edu.au

A. Implementation Details

There are three hyper-parameters β, γ, and α controlling the trade-off of each component to the loss as indicated in Eq. 3 of the manuscript. To verify the effects of changing these parameters on the performance of the proposed method, we conduct extensive experiments by varying these parameters from 0.1 to 0.9 with step size 0.1. The experimental results show that the model achieves the best performance with $\beta = 0.8$, $\gamma = 0.2$, and $\alpha = 0.9$. We didn't use the ImageNet pre-trained model at the SSL stage. The weight decay is set to $5e-4$ and the random seed parameter is set to 1993 for all experiments. Details of other hyper-parameter settings for each dataset are listed in the following.

CUB200, and Mini-ImageNet: At the SSL training stage, we run 100 epochs under a common 5-way 5-shot setting as indicated in Section 4.2 of the manuscript. We use SGD as the optimizer with the initial learning rate being 0.1 and momentum being 0.9. The learning rate drops by a factor of 10 every 40 epochs. The batch size is set to 32. Other settings at the FSCIL stage are identical to CEC.

PlantVillage: The experimental settings at the SSL stage on the PlantVillage dataset are the same as on the CUB and Mini-ImageNet datasets. At the FSCIL stage, we randomly selected 23 classes for base training (session 1). The remaining 15 classes are equally divided into 3 different sessions, thus, each class has 5 images to form a 5-way 5-shot setting. We run 150 epochs with a batch size of 32 and use the SGD optimizer. The initial learning rate is 0.1 and decreases by a factor of 10 at epochs 60 and 90.

CottonCultivar: Because of the lack of training samples, SSL training is performed with a 5-way 3-shot setting on the CottonCultivar dataset. The SSL model is trained for 60 epochs with the learning rate being 0.1 and momentum being 0.9. The learning rate is reduced every 20 epochs by a factor of 10. During the FSCIL stage, half of the classes (40) are used for base training. The other 40 classes are further divided into 8 sessions with 5 images for each class to form a 5-way 5-shot setting. We run 150 epochs with a batch size of 32 and the SGD optimizer. The initial learning rate is set to 0.1 and is decayed by 0.1 at epochs 20, 60, and 90.

SoyCultivarLocal: All the training settings on this dataset are the same as on CottonCultivar except for the data splitting. Since the SoyCultivarLocal dataset has the same total class amount as Mini-ImageNet, we directly adopt its class splitting where 100 classes are used for the base training and the remaining 100 class are equally divided into 10 different sessions, thus each class has 5 images to form a 10-way 5-shot setting.

B. Class Activation Map (CAM) Visual Analysis

To clearly show the effectiveness of the proposed method on different datasets, the class activation map (CAM) is used to demonstrate the detail-enhanced areas of each sample without and with the self-supervised module. Figures 1 - 5 in this supplemental document show the CAM comparisons on five datasets. The top row of each figure represents the original images from the corresponding dataset, the second row is the CAM visualisation of the normal FSCIL method without SSL. The bottom row is the CAM from the proposed model. These CAM maps clearly show that the proposed SSFE-Net is able to locate more precise discriminative features, which means the method enhances the model's ability in discovering more details as well as discriminative areas of images.

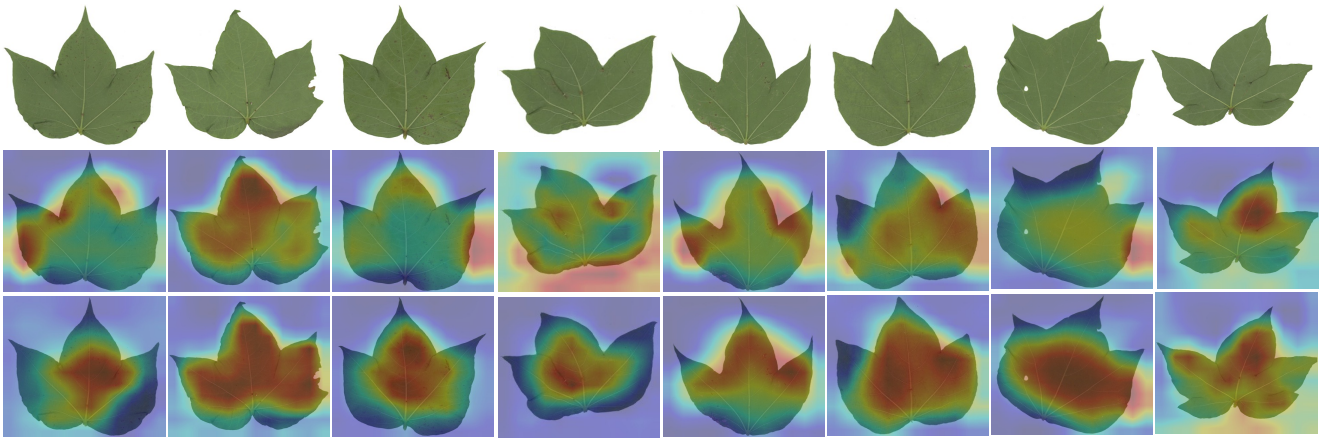

Figure 1: CAM visualisations of eight different samples in CottonCultivar dataset. First row: original; Second row: FSCIL; Third row: SSFE-Net.

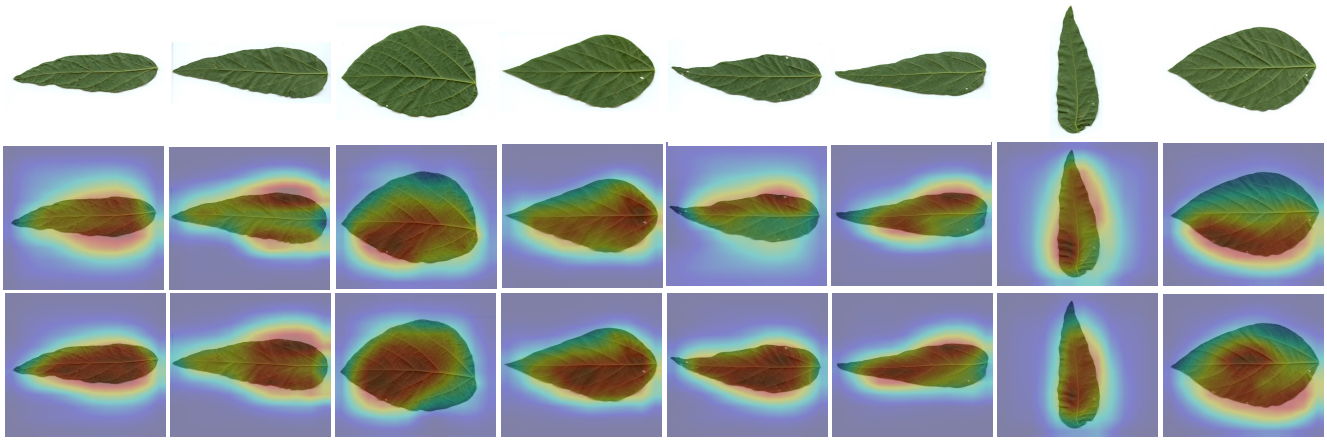

Figure 2: CAM visualisations of eight samples in SoyCultivarLocal dataset. First row: original; Second row: FSCIL; Third row: SSFE-Net.

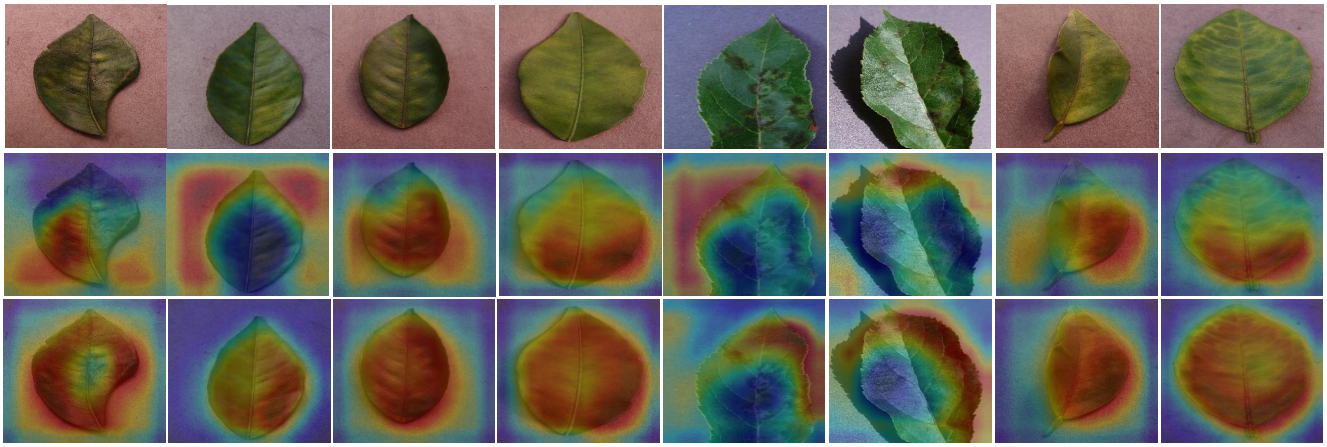

Figure 3: CAM visualisations of eight different samples in PlantVillage dataset. First row: original; Second row: FSCIL; Third row: SSFE-Net.

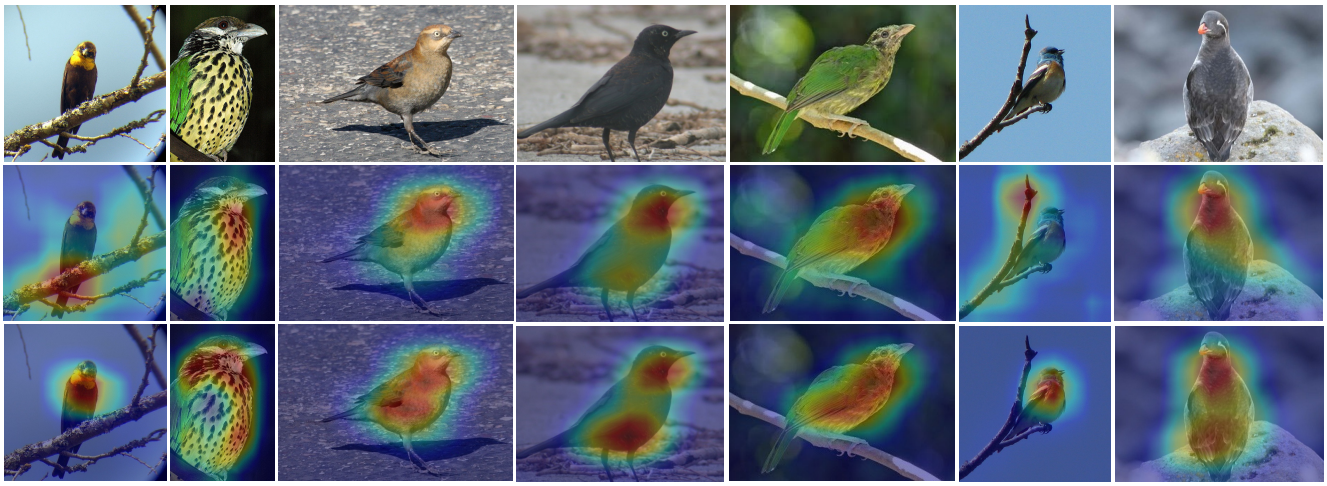

Figure 4: CAM visualisations of seven different samples in CUB200 dataset. First row: original; Second row: FSCIL; Third row: SSFE-Net.

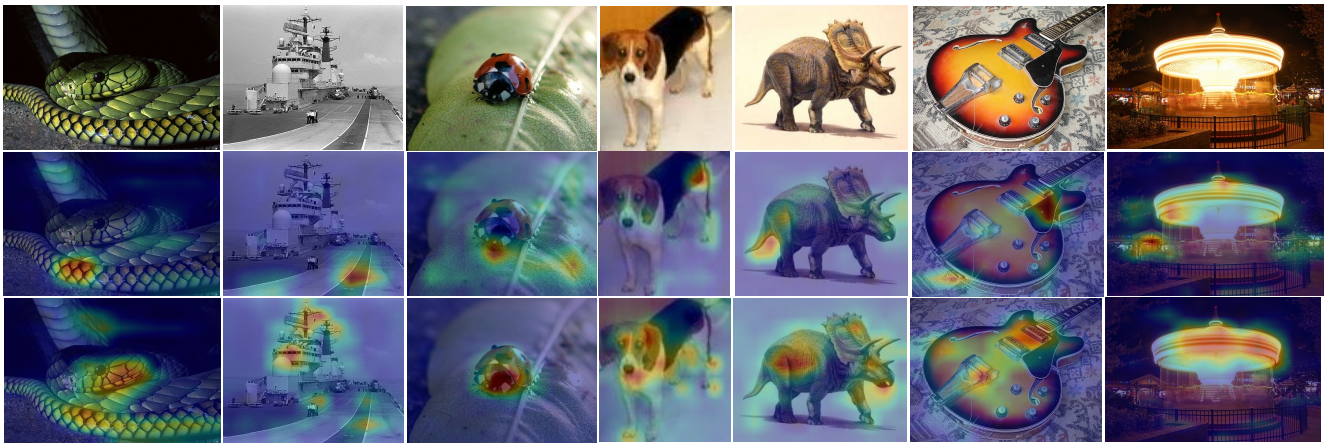

Figure 5: CAM visualisations of seven different samples in Mini-ImageNet dataset. First row: original; Second row: FSCIL; Third row: SSFE-Net.

C. Failure Cases Analysis

Since SSL does not involve the incremental stage of training, the novel class prototypes may suffer from a lack of detail complementing by the SSL model and the samples cannot match the prototypes properly. Figure 6 in the following presents some classification results from different datasets. For each dataset, two images on the same row belong to the same category and the images on the right columns indicate the classification results of that sample. The images surrounded by green boxes indicate the samples successfully match the target classes. Red boxes mark the failure cases of the model. It's clear that the model has great discriminative feature detection ability and is able to distinguish samples with large intra-class and small inter-class variations by analysing their detailed textures. However, when the image background is very complex/cluttered or the samples from the same class deviate too much from others, there is still a chance of failure.

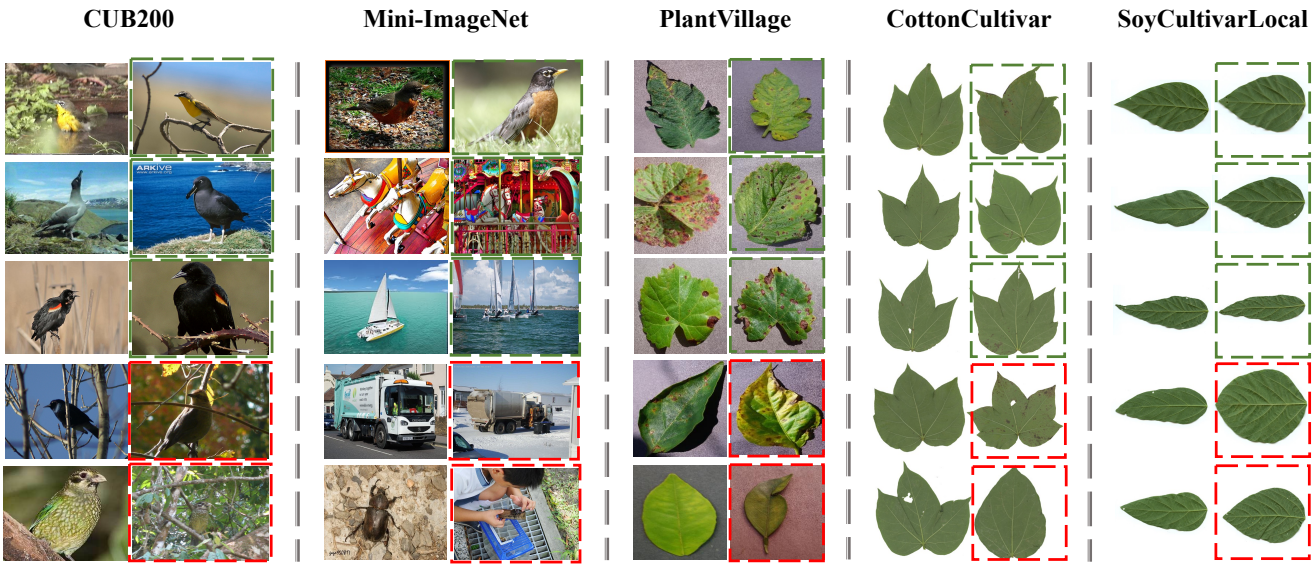

Figure 6: Correctly classified samples (green boxes) and failure cases (red boxes) by the proposed method on different datasets.